

San Francisco California

#### T20

The Unified Medical Language System (UMLS): What Is It and How to Use It?

Olivier Bodenreider, Jan Willis, and William Hole U.S. National Library of Medicine

Wednesday, 8 September 08.30 - 12.00 Union Square 17 & 18



#### Tutorial T20 September 8, 2004



# The Unified Medical Language System What is it and how to use it?





Olivier Bodenreider Jan Willis William Hole

#### Acknowledgements

- ◆ Thanks to NLM Staff and Contractors who contributed to this presentation
  - Kin Wah Fung
  - Betsy Humphreys
  - Rachel Kleinsorge
  - Suresh Srinivasan
  - Karen Thorn
  - Carolyn B. Tilley



#### Outline

- ◆ What is the UMLS?
- ◆ How to use the UMLS?
  - Obtaining a license
  - Remote access
    - Knowledge Source Server (UMLSKS)
    - UMLSKS Application programming interface (API)
  - Local installation and customization (MetamorphoSys)
- **♦** Questions



3

Part I What is the UMLS?

## Outline

- ◆ Part I: What is the UMLS?
  - Introduction
  - Overview through an example
  - The three UMLS Knowledge Sources
    - UMLS Metathesaurus
    - UMLS Semantic Network
    - SPECIALIST Lexicon and lexical tools



5

Part I What is the UMLS?

(1) Introduction

#### What does UMLS stand for?

- **♦** Unified
- ◆ Medical
- **◆** Language
- ◆ System



UMLS®
Unified Medical Language System®
UMLS Metathesaurus®



7

#### Motivation

[Lindberg & al., *Methods*, 1993] [Humphreys & al., *JAMIA*, 1998]

- ♦ Started in 1986
- ◆ National Library of Medicine
- ◆ "Long-term R&D project"
- ◆ Complementary to IAIMS

(Integrated Academic Information Management Systems)

- «[...] the UMLS project is an effort to overcome two significant barriers to effective retrieval of machine-readable information.
- The first is the variety of ways the same concepts are expressed in different machine-readable sources and by different people.
- The second is the distribution of useful information among many disparate databases and systems.»



## The UMLS in practice

- ◆ Database
  - Series of relational files
- **♦** Interfaces
  - Web interface: Knowledge Source Server (UMLSKS)
  - Application programming interfaces (Java and XML-based)
- **♦** Applications
  - lvg (lexical programs)
  - MetamorphoSys (installation and customization)



The UMLS is *not* an end-user application

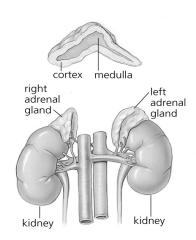
9

# Part I What is the UMLS?

(2) Overview through an example

#### Addison's disease

- Addison's disease is a rare endocrine disorder
- ◆ Addison's disease occurs when the adrenal glands do not produce enough of the hormone cortisol
- For this reason, the disease is sometimes called chronic adrenal insufficiency, or hypocortisolism

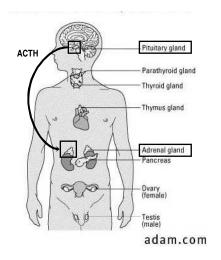




11

#### Adrenal insufficiency Clinical variants

- ◆ Primary / Secondary
  - Primary: lesion of the adrenal glands themselves
  - Secondary: inadequate secretion of ACTH by the pituitary gland
- ◆ Acute / Chronic
- ◆ Isolated / Polyendocrine deficiency syndrome





#### Addison's disease: Symptoms

- ◆ Fatigue
- ♦ Weakness
- ◆ Low blood pressure
- ◆ Pigmentation of the skin (exposed and nonexposed parts of the body)



13

#### AD in medical vocabularies

◆ Synonyms: different terms

Addisonian syndrome

eponym

- Bronzed disease
- Addison melanoderma

symptoms

- Asthenia pigmentosa
- Primary adrenal deficiency

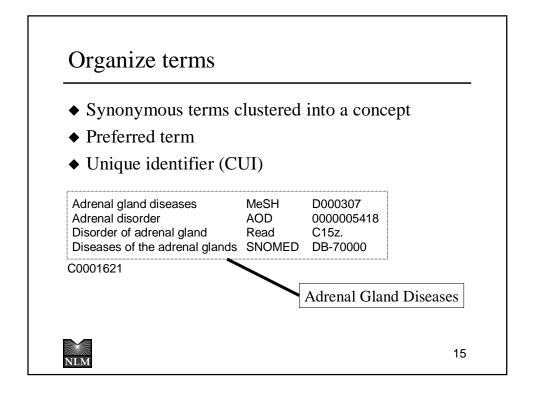
Primary adrenal insufficiency

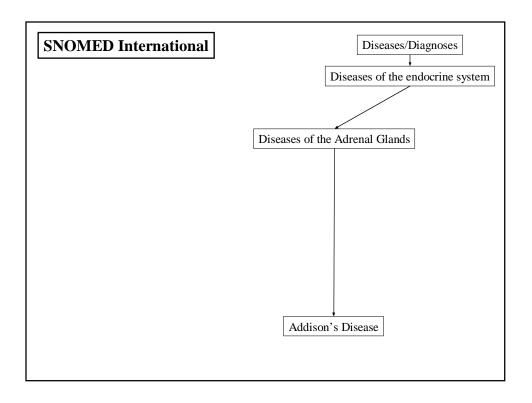
clinical variants

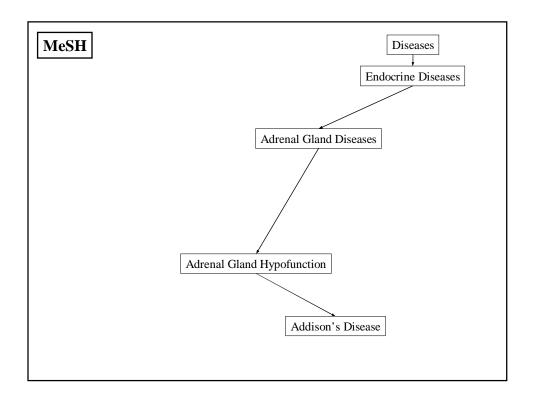
- Primary adrenocortical insufficiency
- Chronic adrenocortical insufficiency

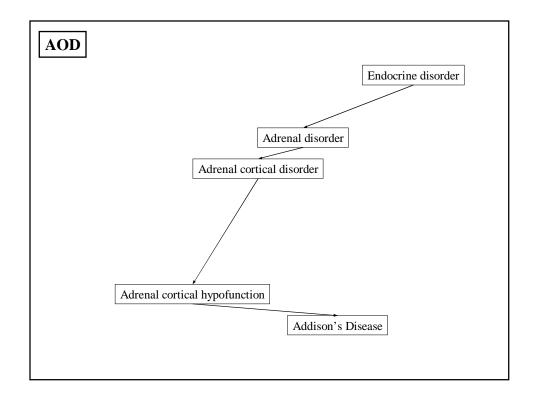
◆ Contexts: different hierarchies

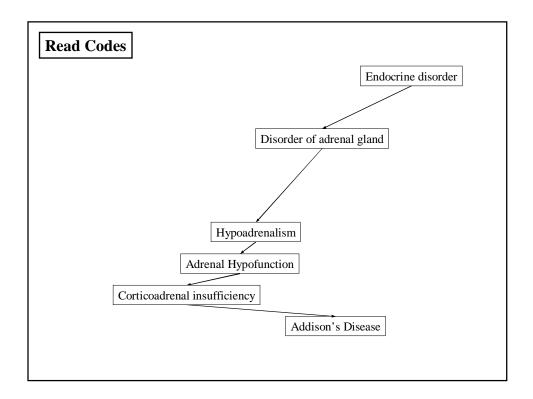


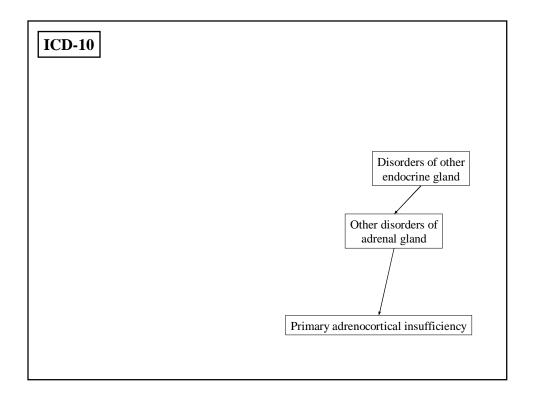






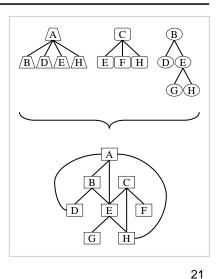






## Organize concepts

- ◆ Inter-concept relationships: hierarchies from the source vocabularies
- Redundancy: multiple paths
- One graph instead of multiple trees (multiple inheritance)





Adrenal Gland Diseases

SNOMED
MeSH
AOD
Read Codes

Hypoadrenalism
Adrenal Gland Hypofunction

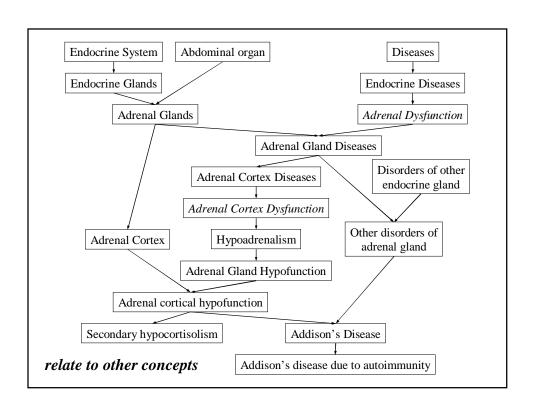
WMLS

Adrenal Cortical hypofunction
Addison's Disease

## Relate to other concepts

- ◆ Additional hierarchical relationships
  - link to other trees
  - make relationships explicit
- ◆ Non-hierarchical relationships
- ◆ Co-occurring concepts
- ◆ Mapping relationships





# Categorize concepts High-level categories (semantic types) Assigned by the Metathesaurus editors Independently of the hierarchies in which these concepts are located Disease or Syndrome Endocrine Diseases Adrenal Gland Diseases Adrenal Gland Hypofunction Addison's Disease

## How do they do that?

- ♦ Lexical knowledge
- ◆ Semantic pre-processing
- **♦** UMLS editors



# Lexical knowledge

Adrenal gland diseases
Adrenal disorder
Disorder of adrenal gland
Diseases of the adrenal glands
C0001621

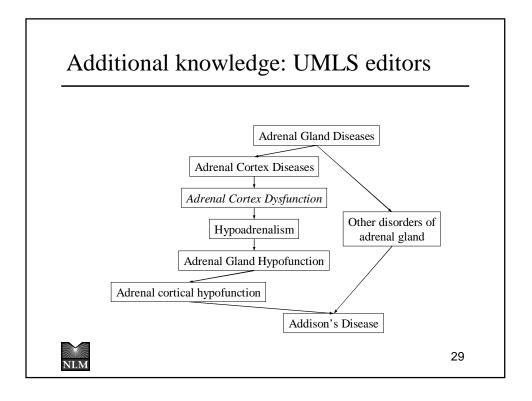


27

# Semantic pre-processing

- ◆ Metadata in the source vocabularies
- ◆ Tentative categorization
- ◆ Positive (or negative) evidence for tentative synonymy relations based on lexical features





#### **UMLS Summary**

- ◆ Synonymous terms clustered into concepts
- ◆ Unique identifier
- ◆ Finer granularity
- ◆ Broader scope
- ◆ Additional hierarchical relationships
- ◆ Semantic categorization



# Part I What is the UMLS?

(3) UMLS Knowledge Sources

# UMLS 3 components

- ◆ Metathesaurus
  - Concepts
  - Inter-concept relationships
- ◆ Semantic Network
  - Semantic types
  - Semantic network relationships
- ◆ Lexical resources
  - SPECIALIST Lexicon
  - Lexical tools



#### **UMLS** Metathesaurus

## Metathesaurus Basic organization

- **◆** Concepts
  - Synonymous terms are clustered into a concept
  - Properties are attached to concepts, e.g.,
    - Unique identifier
    - Definition
- ◆ Relations
  - Concepts are related to other concepts
  - Properties are attached to relations, e.g.,
    - Type of relationship
    - Source



#### Source Vocabularies

(2004AB)

- ◆ 134 source vocabularies
  - 126 contributing concept names
- ◆ 73 families of vocabularies
  - multiple translations (e.g., MeSH, ICPC, ICD-10)
  - variants (American-English equivalents, Australian extension/adaptation)
  - subsequent editions usually considered distinct families (ICD: 9-10; DSM: IIIR-IV)
- Broad coverage of biomedicine
- ◆ Common presentation



35

#### Biomedical terminologies

- ♦ General vocabularies
  - anatomy (UWDA, Neuronames)
  - drugs (RxNorm, First DataBank, Micromedex)
  - medical devices (UMD, SPN)
- Several perspectives
  - clinical terms (SNOMED CT)
  - information sciences (MeSH, CRISP)
  - administrative terminologies (ICD-9-CM, CPT-4)
  - data exchange terminologies (HL7, LOINC)

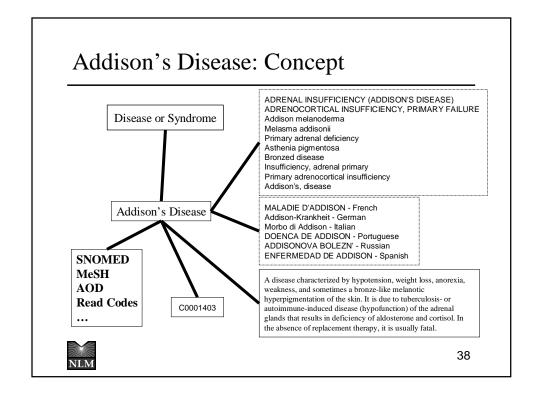


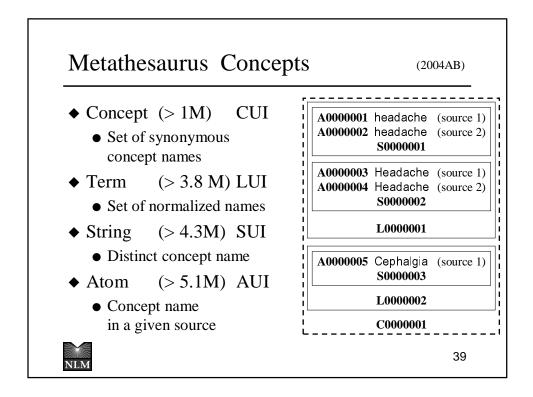
## Biomedical terminologies (cont'd)

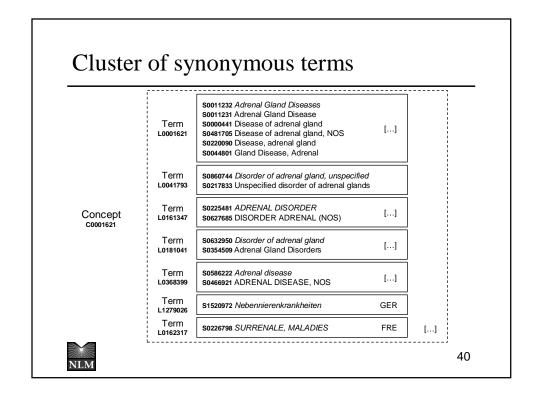
- ◆ Specialized vocabularies
  - nursing (NIC, NOC, NANDA, Omaha, PCDS)
  - dentistry (CDT)
  - oncology (PDQ)
  - psychiatry (DSM, APA)
  - adverse reactions (COSTART, WHO ART)
  - primary care (ICPC)
- ◆ Terminology of knowledge bases (AI/Rheum, DXplain, QMR)



The UMLS serves as a vehicle for the regulatory standards (HIPAA, CHI)

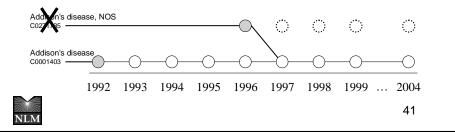






#### Metathesaurus Evolution over time

- ◆ Concepts never die (in principle)
  - CUIs are permanent identifiers
- ◆ What happens when they do die (in reality)?
  - Concepts can merge or split
  - Resulting in new concepts and deletions



#### Metathesaurus Relationships

◆ Symbolic relations: ~9 M pairs of concepts

◆ Statistical relations : ~7 M pairs of concepts (co-occurring concepts)

◆ Mapping relations: 100,000 pairs of concepts

◆ Categorization: Relationships between concepts and semantic types from the Semantic Network



## Symbolic relations

- **♦** Relation
  - Pair of "atom" identifiers
  - Type
  - Attribute (if any)
  - List of sources (for type and attribute)
- ◆ Semantics of the relationship: defined by its type [and attribute]

Source transparency: the information is recorded at the "atom" level



43

## Symbolic relationships Type

◆ Hierarchical

• Parent / Child PAR/CHD

• Broader / Narrower than

RB/RN



Derived from hierarchies

• Possible synonymy

• Siblings (children of parents) SIB

◆ Associative

• Other RO

^ ^

◆ Various flavors of near-synonymy

• Similar RL

• Source asserted synonymy

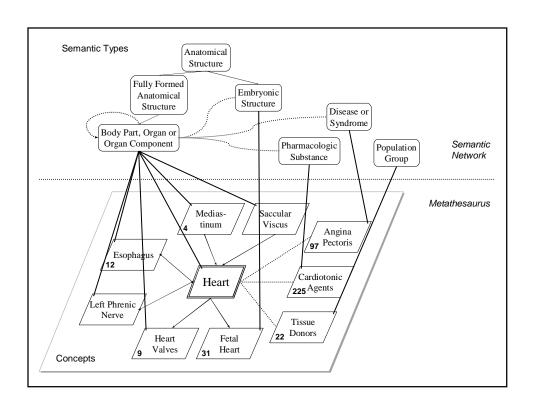
SY RO



# Symbolic relationships Attribute

- ♦ Hierarchical
  - isa (is-a-kind-of)
  - part-of
- ◆ Associative
  - location-of
  - caused-by
  - treats
  - •
- ◆ Cross-references (mapping)





# **UMLS Semantic Network**

## Semantic Network

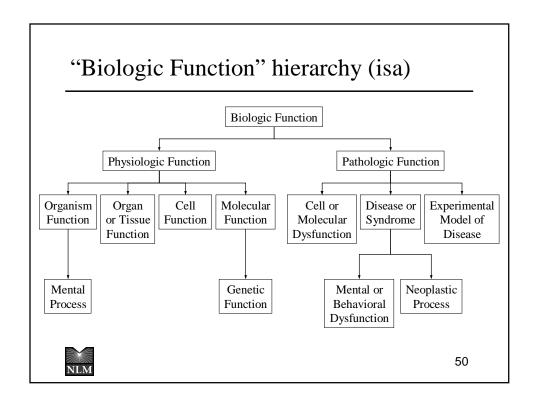
- ◆ Semantic types (135)
  - tree structure
  - 2 major hierarchies
    - Entity
      - Physical Object
      - Conceptual Entity
    - Event
      - Activity
      - Phenomenon or Process

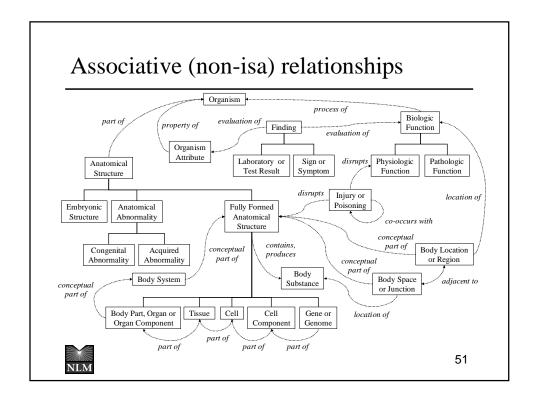


#### Semantic Network

- ◆ Semantic network relationships (54)
  - hierarchical (isa = is a kind of)
    - among types
      - Animal isa Organism
      - Enzyme isa Biologically Active Substance
    - among relations
      - treats isa affects
  - non-hierarchical
    - Sign or Symptom diagnoses Pathologic Function
    - Pharmacologic Substance *treats* Pathologic Function



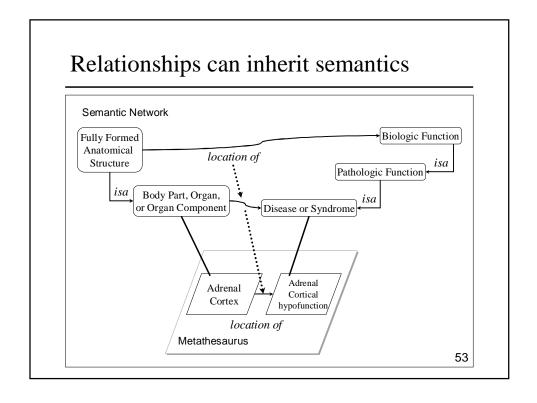




# Why a semantic network?

- ◆ Semantic Types serve as high level categories assigned to Metathesaurus concepts, *independently* of their position in a hierarchy
- ◆ A relationship between 2 Semantic Types (ST) is a possible link between 2 concepts that have been assigned to those STs
  - The relationship may or may not hold at the concept level
  - Other relationships may apply at the concept level





# SPECIALIST Lexicon and lexical tools

#### **SPECIALIST Lexicon**

- **♦** Content
  - English lexicon
  - Many words from the biomedical domain
- ◆ 200,000+ lexical items
- ♦ Word properties
  - morphology
  - orthography
  - syntax
- ◆ Used by the lexical tools



55

## Morphology

- **♦** Inflection
  - noun nucleus, nuclei
  - verb cauterize, cauterizes, cauterized, cauterizing
  - adjective red, redder, reddest
- **♦** Derivation

  - adjective ⟨⇒⟩ noun
     red -- redness



## Orthography

◆ Spelling variants

• oe/e oesophagus - esophagus

• ae/e anaemia - anemia

• ise/ize cauterise - cauterize

• genitive mark Addison's disease

Addison disease Addisons disease



57

#### **Syntax**

- **♦** Complementation
  - verbs

■ intransitive I'll treat.

transitive He treated the patient.

ditransitive
He treated the patient with a drug.

- nouns
  - prepositional phrase

Valve of coronary sinus

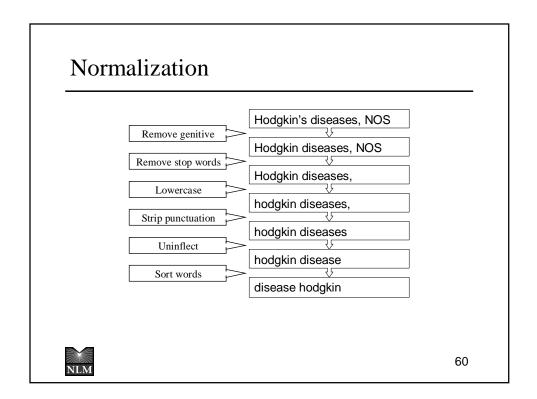
◆ Position for adjectives

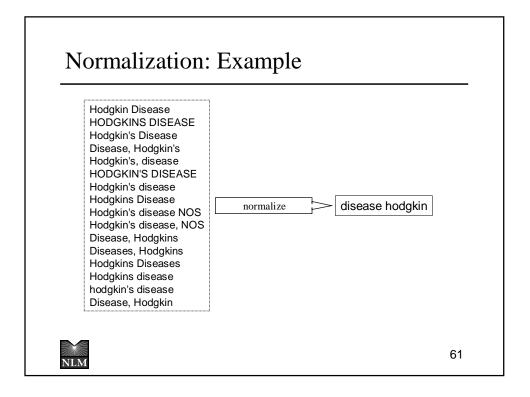


#### Lexical tools

- ◆ To manage lexical variation in biomedical terminologies
- ◆ Major tools
  - Normalization
  - Indexes
  - Lexical Variant Generation program (lvg)
- ◆ Based on the SPECIALIST Lexicon
- ◆ Used by noun phrase extractors, search engines







#### Normalization Applications

- ◆ Model for lexical resemblance
- ◆ Help find lexical variants for a term
  - Terms that normalize the same usually share the same LUI
- ◆ Help find candidates to synonymy among terms
- ◆ Help map input terms to UMLS concepts



#### Indexes

- ♦ Word index
  - word to Metathesaurus strings
  - one word index per language
- ◆ Normalized word index
  - normalized word to Metathesaurus strings
  - English only
- ◆ Normalized string index
  - normalized term to Metathesaurus strings
  - English only



63

#### Lexical Variant Generation program

- ◆ Tool for specialists (linguists)
- ◆ Performs atomic lexical transformations
  - generating inflectional variants
  - lowercase
  - ...
- ◆ Performs sequences of atomic transformations
  - a specialized sequence of transformations provides the normalized form of a term (the *norm* program)



# Part II How to use the UMLS?

## Outline

- ◆ Part II: *How to use the UMLS?* 
  - Obtaining a license
  - Remote access
    - Knowledge Source Server (UMLSKS)
    - UMLSKS Application programming interface (API)
  - Local installation and customization (MetamorphoSys)



# Part II How to use the UMLS?

(1) Obtaining a license

#### First step License agreement

♦ Online Web-based license:

http://www.nlm.nih.gov/research/umls/license.html

• Read license

Accept and continue

• Read appendix

Accept

• Print a copy for your records

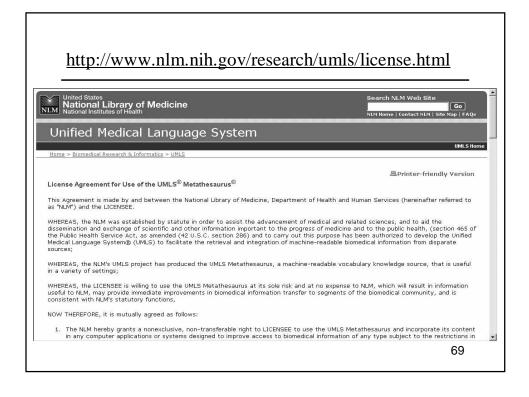
Printer-friendly Version

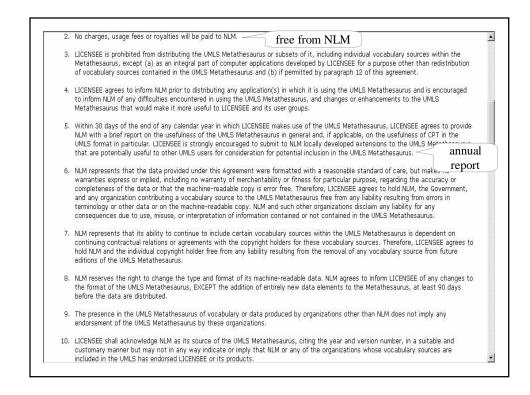
• Complete the Web form

Submit

- Verify:
  - receive e-mail from NLM; go to Web site within 72 hours and enter first and last name
- NLM official will countersign (turn-around time of a few days)
- Receive 2<sup>nd</sup> e-mail from NLM with new license number





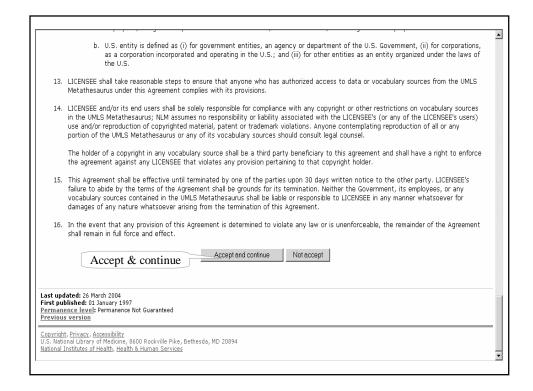


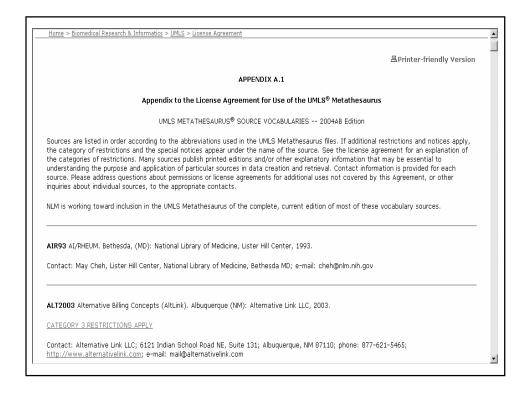
Some of the Material in the UMLS Metathesaurus is from copyrighted sources. If LICENSEE uses any material from copyrighted sources from the UMLS Metathesaurus a) the LICENSEE is required to display in full, prior to providing user access to the Metathesaurus or any of the vocabulary sources within the UMLS, the following wording in order that its users be made aware of these copyright constraints: Some material in the UMLS Metathesaurus is from copyrighted sources of the respective copyright holders. Users of the UMLS Metathesaurus are solely responsible for compliance with any copyright, patent or trademark restrictions and are referred to the copyright, patent or trademark notices appearing in the original sources, all of which are hereby incorporated by reference."; to display a list of all of the vocabularies contained within the UMLS Metathesaurus that are used in the LICENSEE's application; and to indicate for each vocabulary any appropriate copyright notice and whether the entire contents is present or only a portion of it b) the LICENSEE is prohibited from altering UMLS and other vocabulary source content contained within the UMLS Metathesaurus, but may include content from other sources in applications that also contain content from the UMLS Metathesaurus. The LICENSEE may not imply in any way that data from other sources is part of the UMLS Metathesaurus or of any of its vocabulary sources. c) the LICENSEE is required to include in its applications identifiers from the UMLS Metathesaurus such that the original source vocabularies for any data obtained from the UMLS Metathesaurus can be determined by reference to a complete version of the UMLS Metathesaurus. need to retain identifiers 12. For material in the UMLS Metathesaurus obtained from some sources additional restrictions on LICENSEE's use may apply. The categories of additional restrictions are described below. The list of UMLS Metathesaurus Vocabulary Sources, which is part of this Agreement and is updated when each version of the Metathesaurus is released, indicates the category of additional restrictions, if any, that apply to each vocabulary source. additional licenses may be necessary LICENSEE should contact the copyright holder directly to discuss uses of a source vocabulary beyond those allowed under this license agreement. If LICENSEE or LICENSEE's end user has a separate agreement with the copyright holder for use of a UMLS Metathesaurus source vocabulary, LICENSEE or LICENSEE's end user may use vocabulary source content obtained from the UMLS Metathesaurus in accordance with the terms of the separate agreement. 12. 1. Category 1: LICENSEE is prohibited from translating the vocabulary source into another language or from producing other derivative works based on this single vocabulary source

#### • 12. 2. Category 2: All category 1 restrictions AND LICENSEE is prohibited from using the vocabulary source in operational applications that create records or information containing data from the vocabulary source. Use for data creation research or product development is allowed. 12. 3. Category 3: LICENSEE's right to use material from the source vocabulary is restricted to internal use at the LICENSEE's site(s) for research, product development, and statistical analysis only. Internal use includes use by employees, faculty, and students of a single institution at multiple sites. Notwithstanding the foregoing, use by students is limited to doing research under the direct supervision of faculty Internal research, product development, and statistical analysis use expressly excludes: use of material from these copyrighted sources in routine patient data creation; incorporation of material from these copyrighted sources in any publicly accessible computer-based information system or public electronic bulletin board including the Internet; publishing or translating or creating derivative works from material from these copyrighted sources; selling, leasing, licensing, or otherwise making available material from these copyrighted works to any unauthorized party; and copying for any purpose except for back up or archival purposes. LICENSEE may be required to display special copyright, patent and/or trademark notices before displaying content from the vocabulary source. Applicable notices are included in the list of UMLS Metathesaurus Vocabulary sources, that is part of this Agreement. new SNOMED category 12. 4. Category 4: 12.4.1. LICENSEE is prohibited from translating the vocabulary source into another language or from altering the vocabulary 12.4.2. LICENSEE's right to use the vocabulary source is restricted to use in the U.S. by LICENSEE's employees, contractors, faculty, students, clients, patients, or constituents within electronic systems or devices built, purchased, licensed, or used by LICENSEE for U.S. governmental purposes or for any health care, public health, research, educational, or statistical use in the U.S. Use by students is limited to research or educational activities under the direct supervision of faculty 12.4.3. LICENSEE has the right to distribute the vocabulary source in the U.S., but only in combination with other UMLS Metathesaurus content, Further, LICENSEE's right to distribute is restricted to a. Electronic distribution to LICENSEE's direct U.S. affiliates, or to other U.S. entities that have signed the UMLS license, in

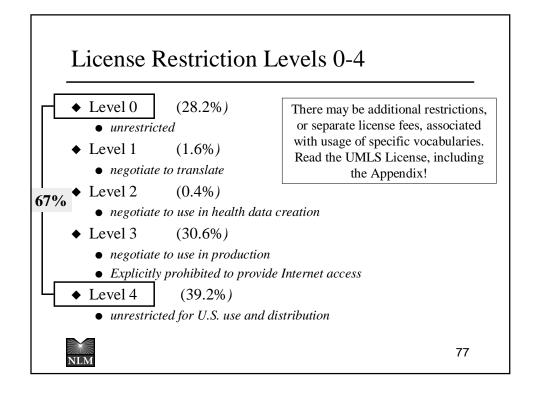
order to facilitate use of the vocabulary for health care, public health, research, educational or statistical purposes in the

#### U.S. only i. LICENSEE must take reasonable precautions to prevent distribution of the vocabulary source to non-US entities. ii. LICENSEE must include in its annual report a list of all U.S. affiliates or other U.S. entities to whom it has distributed content from the vocabulary source b. Distribution of encoded patient level data sets or knowledge encoded in the vocabulary source by LICENSEE to any U.S. Inclusion of encoded records or content from the vocabulary source in: (1) free publicly accessible retrieval systems or (2) fee-based retrieval systems that are accessible within the U.S. only, provided that these systems do not permit users to copy or extract any significant portion of the vocabulary source. a. U.S. is defined as all U.S. states, territories, and the District of Columbia; any U.S. government facility or office, whether permanent or temporary, wherever located; and access to a system in any of these locations by U.S. government employees, designated representatives or contractors, wherever located, for U.S. government purposes. b. U.S. entity is defined as (i) for government entities, an agency or department of the U.S. Government, (ii) for corporations, as a corporation incorporated and operating in the U.S.; and (iii) for other entities as an entity organized under the laws of 13. LICENSEE shall take reasonable steps to ensure that anyone who has authorized access to data or vocabulary sources from the UMLS Metathesaurus under this Agreement complies with its provisions. 14. LICENSEE and/or its end users shall be solely responsible for compliance with any copyright or other restrictions on vocabulary sources in the UMLS Metathesaurus; NLM assumes no responsibility or liability associated with the LICENSEE's (or any of the LICENSEE's users) use and/or reproduction of copyrighted material, patent or trademark violations. Anyone contemplating reproduction of all or any portion of the UMLS Metathesaurus or any of its vocabulary sources should consult legal counsel. The holder of a copyright in any vocabulary source shall be a third party beneficiary to this agreement and shall have a right to enforce the agreement against any LICENSEE that violates any provision pertaining to that copyright holder. This Agreement shall be effective until terminated by one of the parties upon 30 days written notice to the other party. LICENSEE's failure to abide by the terms of the Agreement shall be grounds for its termination. Neither the Government, its employees, or any









Part II How to use the UMLS?

(2) Remote access

#### Remote Access

- ◆ UMLS Knowledge Source Server: http://umlsks.nlm.nih.gov
- ◆ Web search interface
- ◆ Application Programming Interface (API)



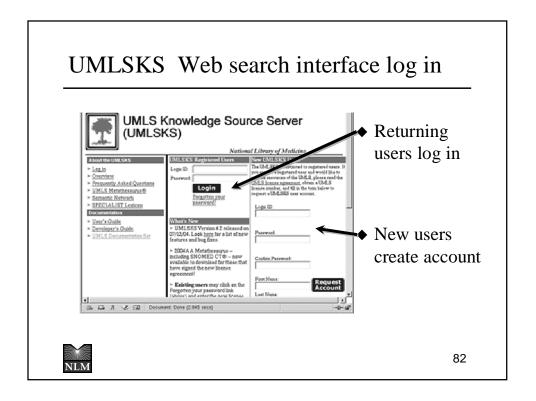
79

Knowledge Source Server Web search interface

#### UMLSKS Web search interface

- ◆ Logging in
- ♦ Basic searching
- ◆ Advanced searching

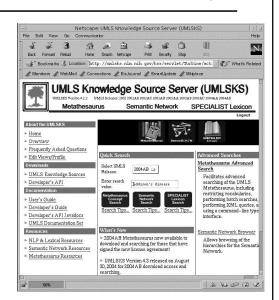


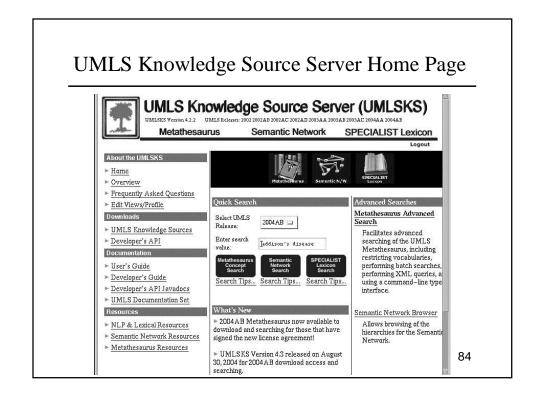


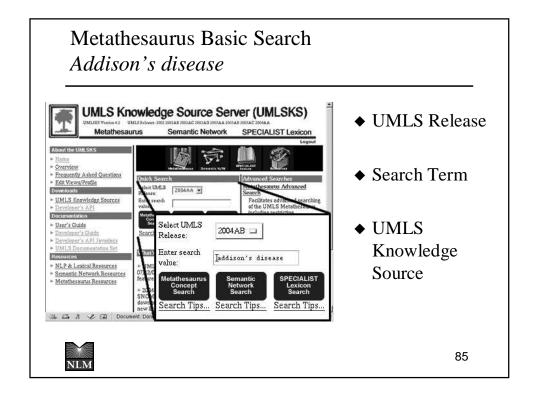
#### UMLS Knowledge Source Server Home Page

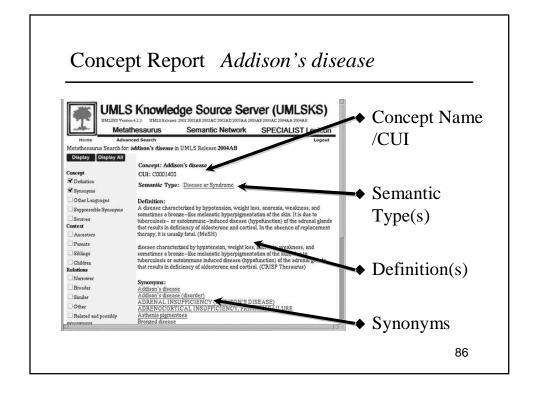
- ◆ Tabs across top access basic searching of 3 Knowledge Sources
- Advanced searching options on right-hand side

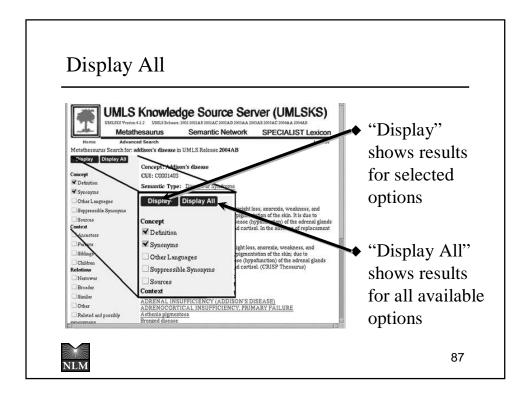


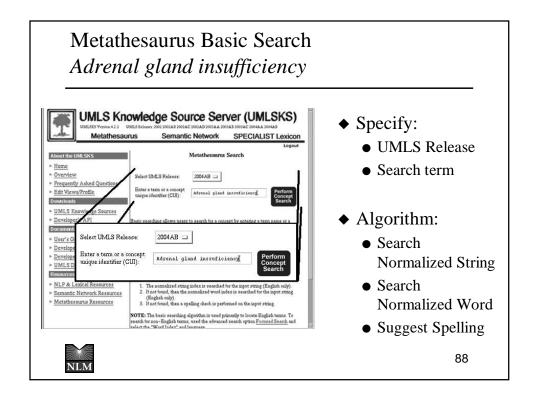


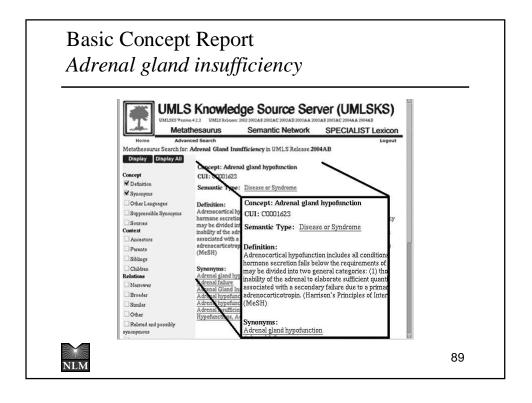


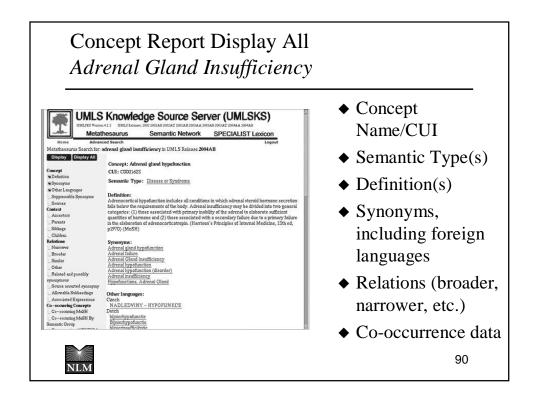


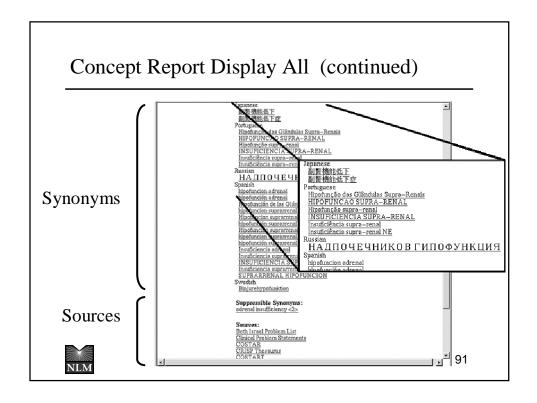


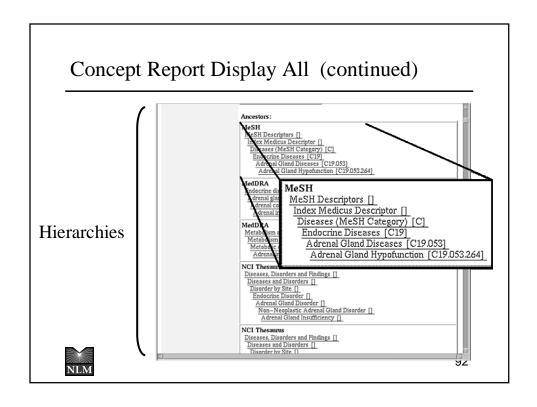


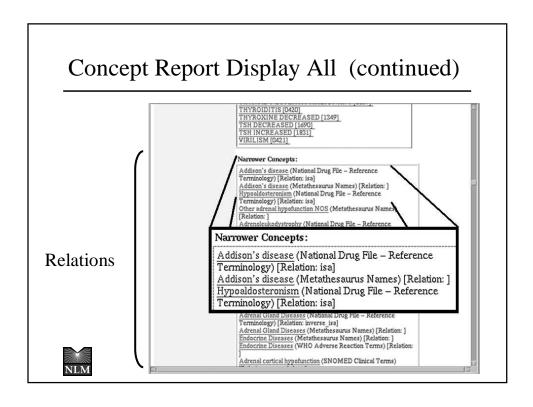


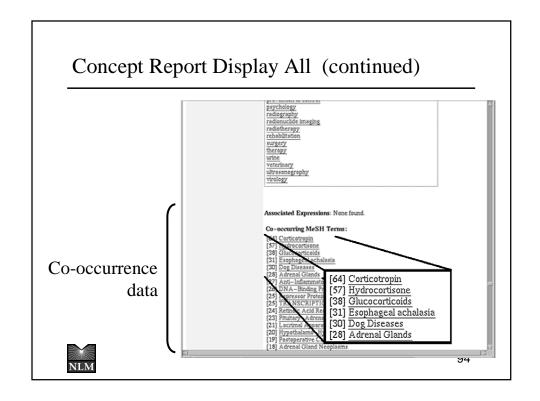






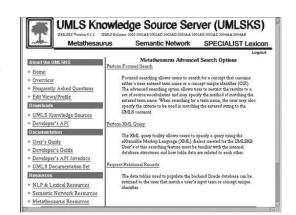






#### Metathesaurus Advanced Search Options

- ♦ Focused Search
- ◆ Raw Relational Records



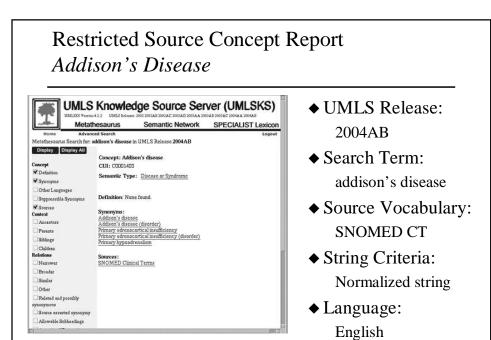


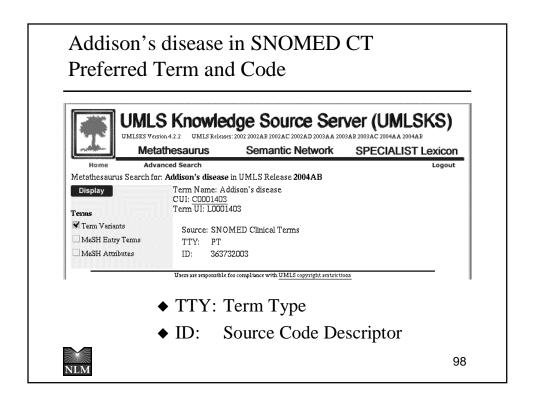
95

# Metathesaurus Advanced Search Feature Focused Search



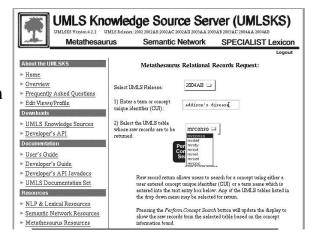
- **♦** UMLS Release
- ◆ Search Term
- ◆ Source Vocabularies
- ◆ String Criteria
  - Exact Match
  - Normalized string & word
  - Word
  - Truncation (left/right)
  - Approximate Match
- ◆ Language





#### Metathesaurus Advanced Search Feature Relational Record Request

- ◆ UMLS Release
- ◆ Search Term
- ◆ UMLS
  Relational
  Table





99

#### Relational Records MRCONSO.RRF

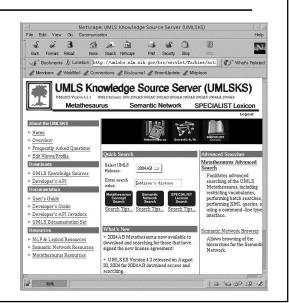
CUI|LAT|TS|LUI|STT|SUI|ISPREF|AUI|SAUI|SCUI|SDUI|SAB|TTY|CODE|STR|SUPPRESS|CWF|
C0001403|CZE|P|L3180742|PF|S3708232|Y|A3910108||1D000224|MSHCZE|MH|D000224|ADDISONOWA NEMOC|
C0001403|DUT|P|L2048638|PF|S2386860|N|A5566810||110001130|MSRDUIT|LT|10001130|Addison, ziekte v
C0001403|DUT|P|L2048638|PF|S2386865|Y|A3931188||1D000224|MSHDUT|SY|D000224|Addison, ziekte v
C0001403|DUT|S|L2048637|PF|S2386859|Y|A3931188||1D000224|MSHDUT|SY|D000224|Addison, syndrom
C0001403|DUT|S|L2528364|PF|S2386859|Y|A3931188||1D000224|MSHDUT|SY|D000224|Addison, syndrom
C0001403|DUT|S|L2528364|PF|S2386531|Y|A396882||1D000224|MSHDUT|SY|D000224|Ziekte van Addisc
C0001403|DUT|S|L3205108|PF|S3732602|Y|A3970882||1D000224|MSHDUT|SY|D000224|Primaire bijniers
C0001403|DUT|S|L3999230|PF|S5886775|Y|A3970882||1D000224|MSHDUT|SY|D000224|Primaire bijnier
C0001403|DUT|S|L3999230|PF|S5886775|Y|A6627443||110052381|MDRDUT|LT|10052381|primaire bijnie
C0001403|DUT|S|L3999270|PF|S5886775|Y|A6627494||110052381|MDRDUT|PT|10052381|primaire bijnie
C0001403|DUT|S|L3999270|PF|S5886775|Y|A6627494||110052381|MDRDUT|PT|10052381|primaire bijnie
C0001403|DUT|S|L3999270|PF|S5886775|Y|A6627494||10052381|MDRDUT|PT|10052381|primaire bijnie
C0001403|DUT|S|L3999270|PF|S5886775|Y|A6627494||10052381|MDRDUT|PT|10052381|primaire bijnie
C0001403|DUT|S|L3999270|PF|S0354372|N|A0388278||1|C0001403|PG|D000006012|Addison's disease|01|
C0001403|ENG|P|L0001403|PF|S0354372|N|A03882879||1|C0F|PT|0000061|Addison's disease|01|
C0001403|ENG|P|L0001403|PF|S0354372|N|A0388280||10000006012|Addison's disease|01|
C0001403|ENG|P|L0001403|PF|S0354372|N|A0388280||110001390|MDR|LT|1001310|Addison's disease|
C0001403|ENG|P|L0001403|PF|S0354372|N|A0388280||110001390|MDR|LT|1000130|Addison's disease|
C0001403|ENG|P|L0001403|PF|S0354372|N|A0388280||110001390|MDR|LT|1000130|Addison's disease|



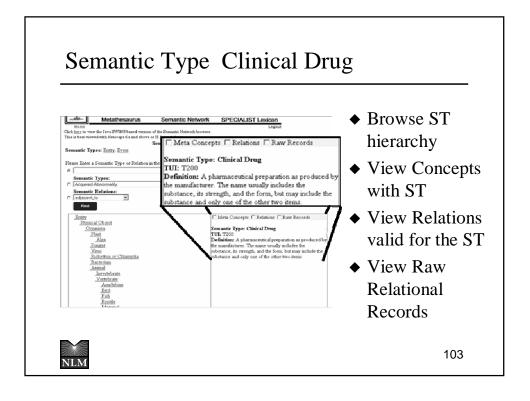
## Semantic Network Searching

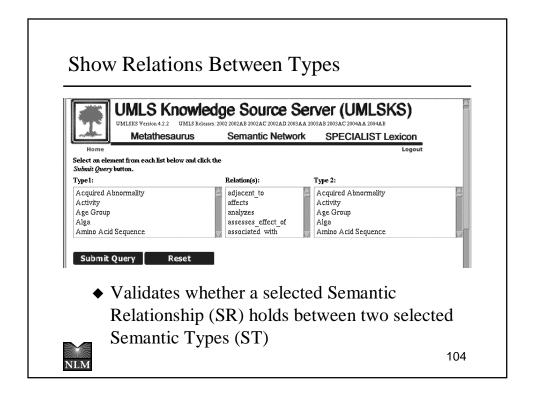
- Select Tab along top
- ♦ Quick search
- ◆ Advanced Search on right-hand side

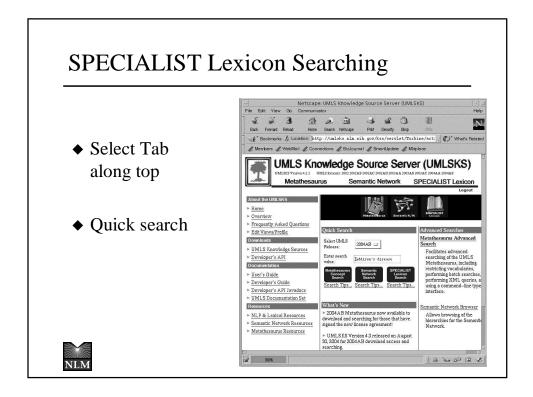


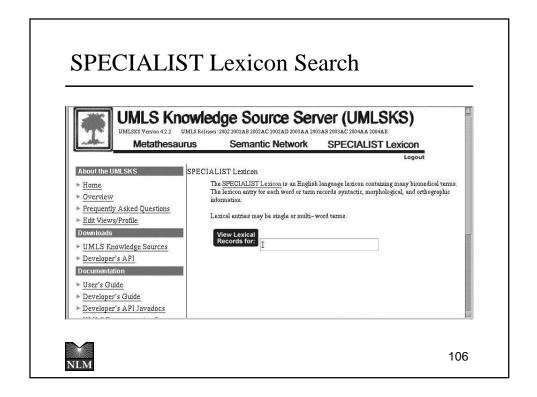


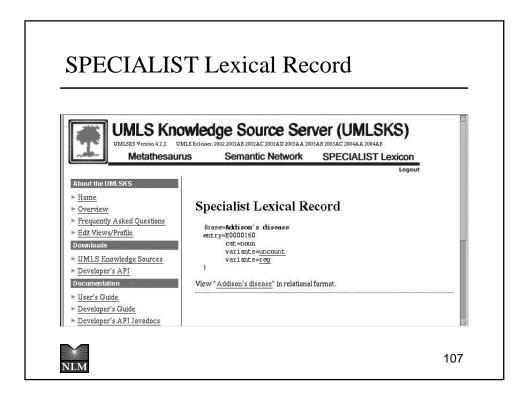
#### Semantic Network Search ♦ Enter search UMLS Knowledge Source Server (UMLSKS) string Semantic Network SPECIALIST Lexicon Metathesaurus Mome Click <u>here</u> to view the Java SWING based version of the Semantic Network browser. This is best viewed with Netscape 6.x and above or IE 5.x and above. -or-Semantic Network Semantic Types: Entity, Event ◆ Select semantic Please Enter a Semantic Type or Relation in the box below or select from the list below and click on the Find button: type Acquired Abnormality -or-Semantic Relations: ◆ Select semantic relation 102

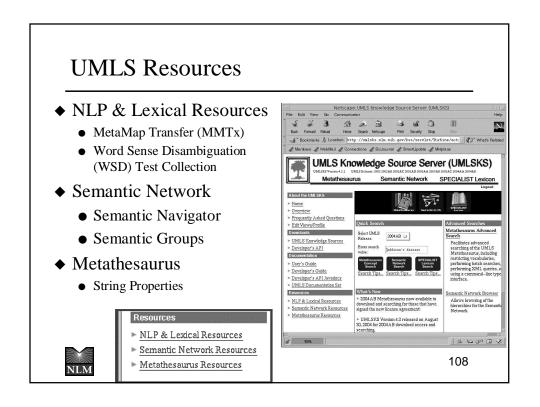


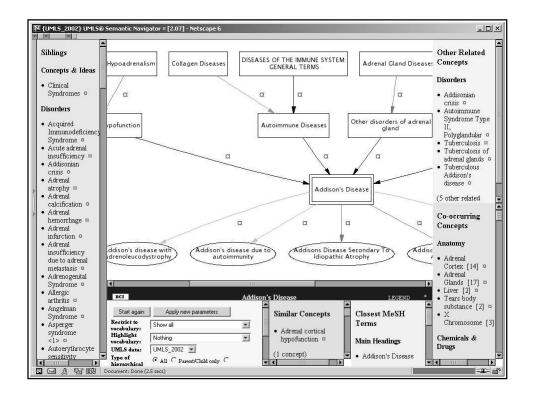












Knowledge Source Server

Application Programming Interface

#### **UMLSKS** API basics

- ◆ Remote server at NLM
- ◆ Local application connected through

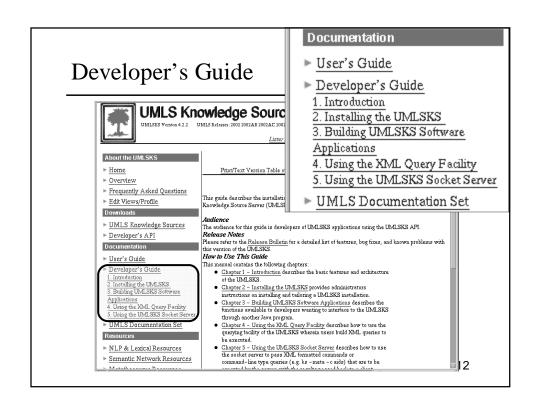
#### Java RMI

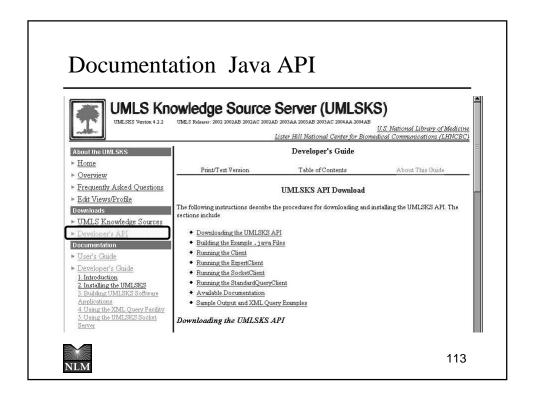
- ◆ Java-based applications
- ◆ Developer's Guide: Chapter 3
- Set of Java classes (part of the UMLSKS API download)
- Detailed Javadoc documentation online and with API download

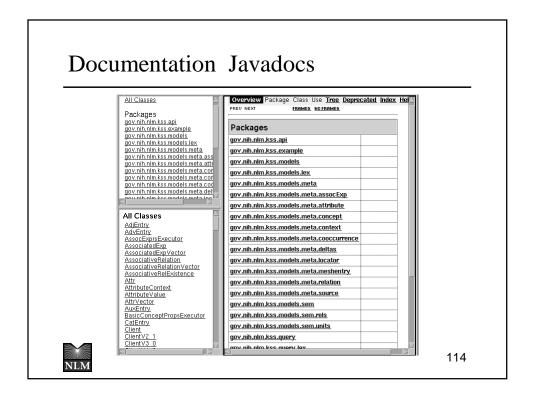
#### TCP/IP socket

- XML-based queries
- ◆ Developer's Guide: Chapter 5
- ♦ XML schema
- Socket server
  - Host: umlsks.nlm.nih.gov
  - Port: 8042









## Sample XML query (1) Current version

```
<?xml version="1.0"?>
<getCurrentUMLSVersion version="1.0"/>
```



<?xml version="1.0"?> <CurrentUMLSYear version="1.0"> 2004AB </CurrentUMLSYear>



115

#### Sample XML query (2) Concepts by string

```
<?xml version="1.0"?>
<findCUI version="1.0">
<conceptName>appendicectomy</conceptName>
<language>ENG</language>
<exact/>
<noSuppressibles/>
</findCUI>
```





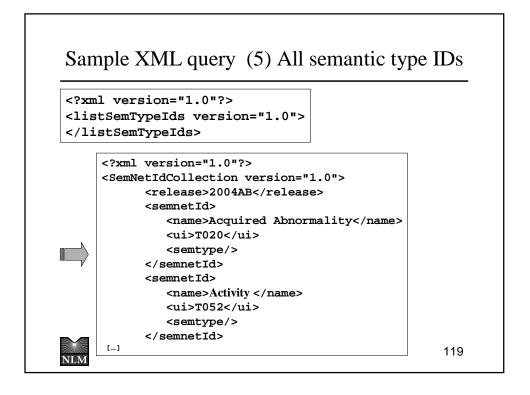
<?xml version="1.0"?> <ConceptIdCollection version="1.0"> <release>2004AB</release> <conceptId> <cui>C0003611</cui> <cn>Appendectomy</cn> </conceptId> </ConceptIdCollection>

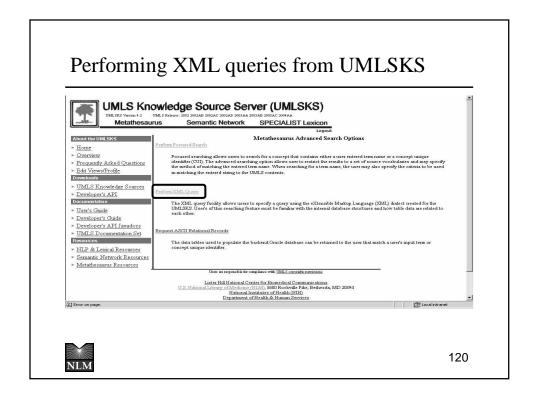
117

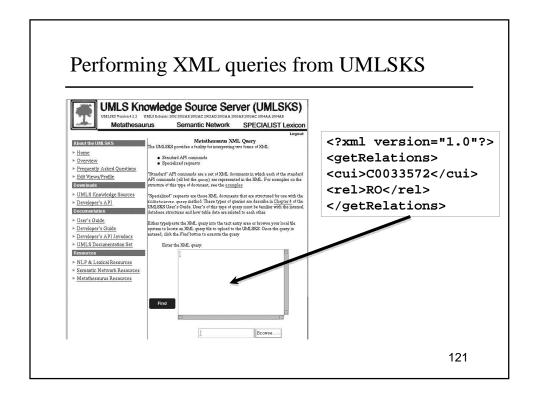
#### 

</SemanticTypeCollection>

#### Sample XML query (4) Relationships <?xml version="1.0"?> <getRelations version="1.0"> <cui>C0033572</cui> <rel>RO</rel> </getRelations> <?xml version="1.0"?> <RelationCollection version="1.0"> <relation> <rel>RO</rel> <cui2>C0005001</cui2> <cn2>Benign prostatic hyperplasia</cn2> <rela>has\_finding\_site</rela> <sab>SNOMEDCT</sab> <sl>SNOMEDCT</sl> </relation> <del>1</del>18







# Part II How to use the UMLS?

(3) Installing the UMLS locally and Customizing the Metathesaurus using MetamorphoSys

## What is MetamorphoSys?

- ◆ Tool distributed with the UMLS
- ◆ Multi-platform Java software
- ◆ The UMLS installation and customization wizard
  - Installs Knowledge Sources to local storage
  - Subsets and customizes a local Metathesaurus



123

## Using MetamorphoSys

- ◆ Simple to use
- ◆ Screens and tabs lead you through process
- ◆ Installs NLM data format files to local storage



## Why use MetamorphoSys?

Customize the Metathesaurus

- ◆ To remove terminology that is unhelpful, or even harmful, to your needs and purposes
- ◆ To comply with terms of license agreement



125

## Why use MetamorphoSys?

Changing Default Settings

- ◆ To alter the preferred name
- ◆ To alter suppressibility of specific source term types



#### **Customization is Critical**

- Requires a clear understanding of:
  - Characteristics of source vocabularies
  - License arrangements
  - User's functional requirements
  - User's purpose and perspective
- ◆ Technical expertise

## ... and requires a multidisciplinary technical team

NLM

127

#### Machine Requirements

- ◆ A fast CPU 1 GHz or higher
- ◆ 1 GB RAM recommended (512 MB min.)
- ♦ 6x (or better) DVD drive
- ◆ 22 GB minimum free disk space
- ◆ Runs on Sun Solaris 8 & 9, Windows XP, NT, and 2000, Linux, and Mac
- ◆ 1-10 hours run time on platforms tested



#### Download from UMLSKS ...

- ◆ High speed Internet connection required
- Read the README file for the release
- ◆ 2004AB UMLS Files

2004AB.CHK

2004AB.MD5

2004ab-1-meta.nlm

2004ab-2-meta.nlm

2004ab-3-meta.nlm

mmsys.zip

Copyright\_Notice.txt

README.txt -



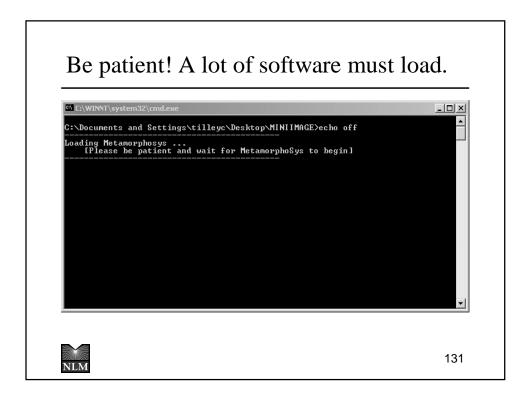
Please README!

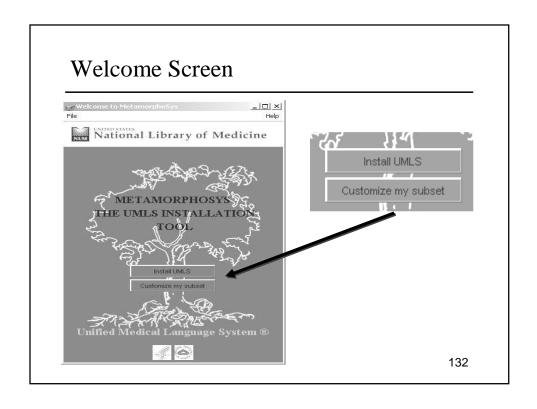
129

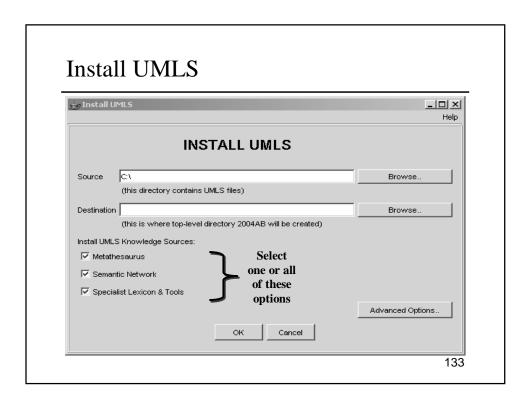
#### ...or DVD?

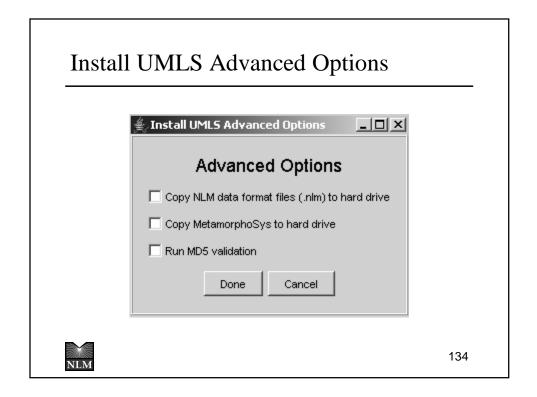
- ◆ Order at: <u>umls\_support@nlm.nih.gov</u>
- **◆ Include your license number**
- ◆ Run MetamorphoSys from DVD
  - Windows
    - Autorun; or go to root directory and click on "windows\_mmsys.bat"
  - Linux, Solaris, Macintosh
    - open a terminal window, change to the root directory and type appropriate command: ./linux\_mmsys.sh, ./solaris\_mmsys.sh, ./macintosh\_mmsys.sh

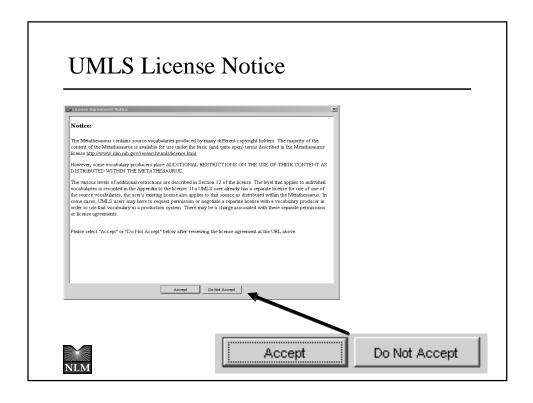


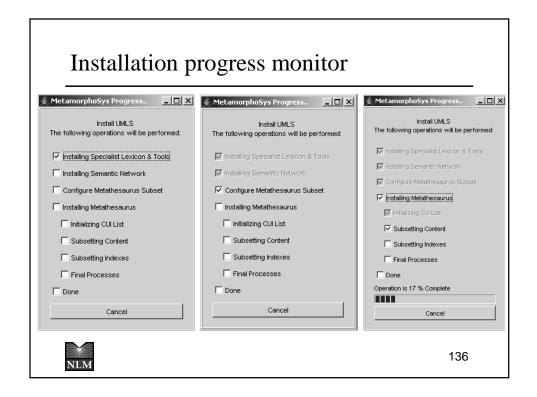


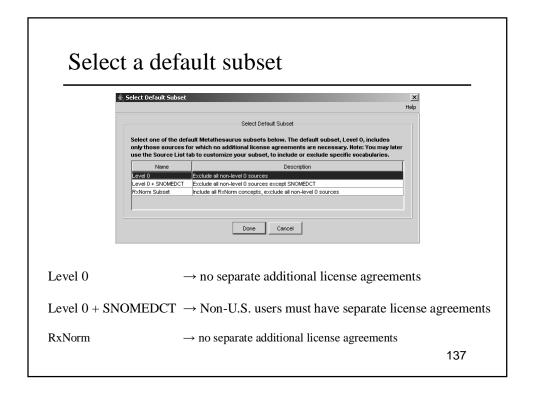


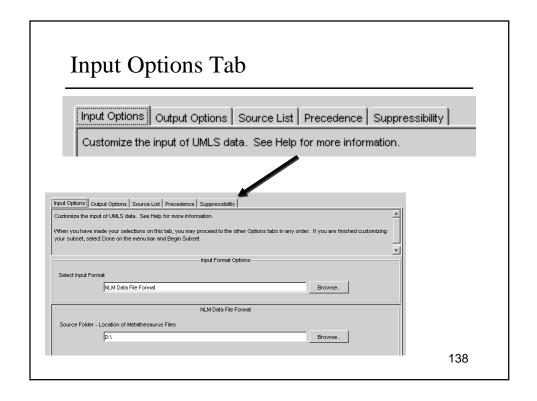


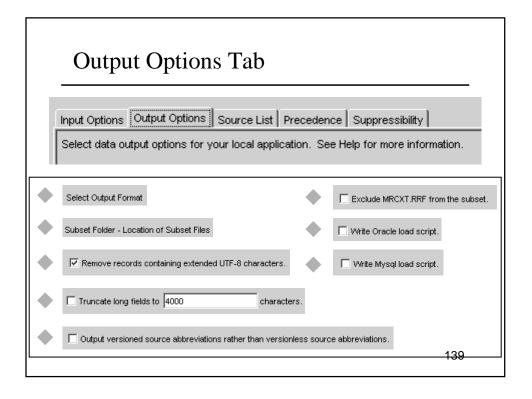


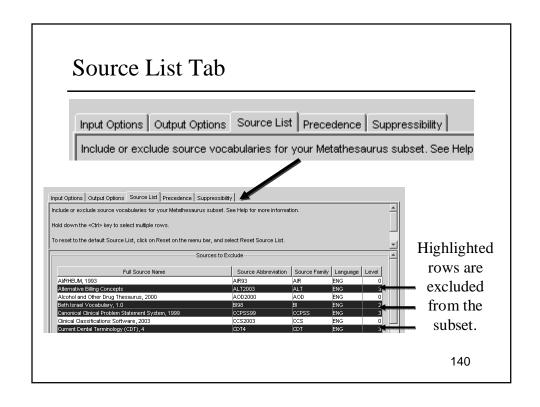


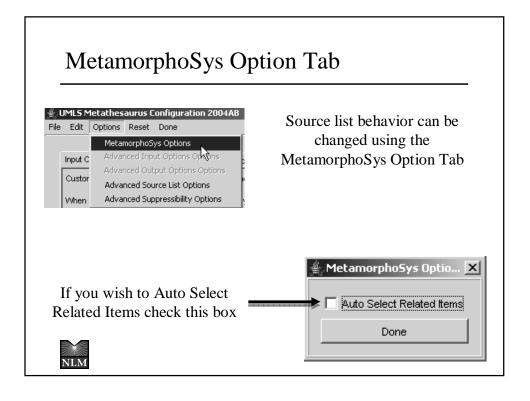


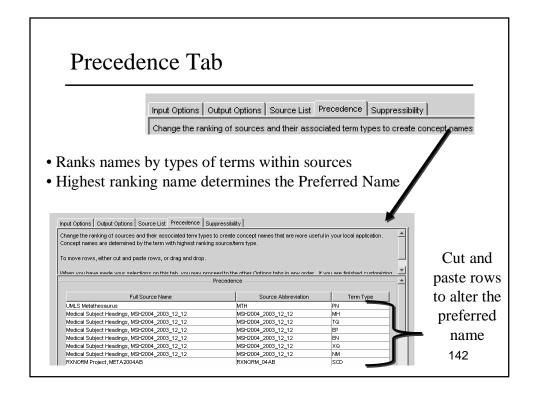


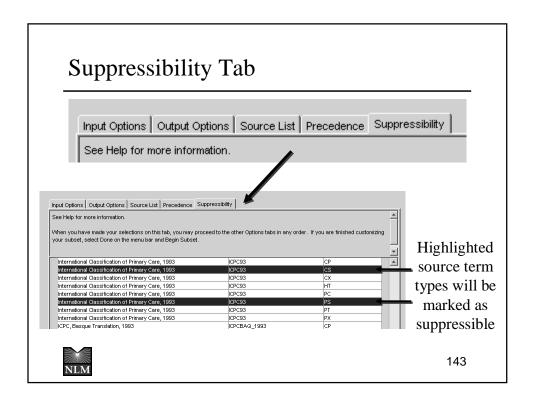


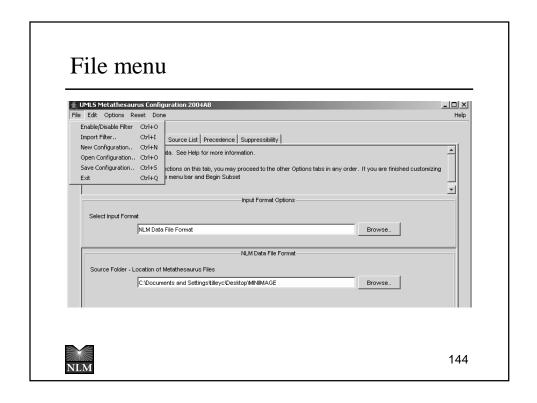


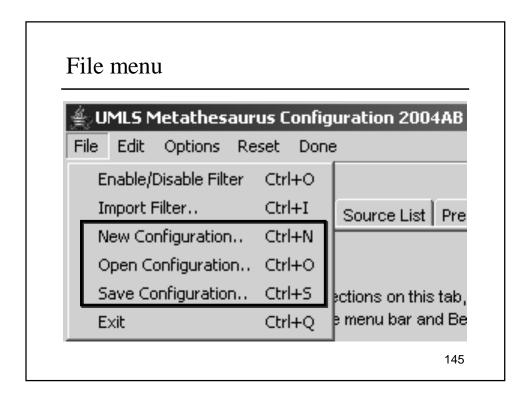


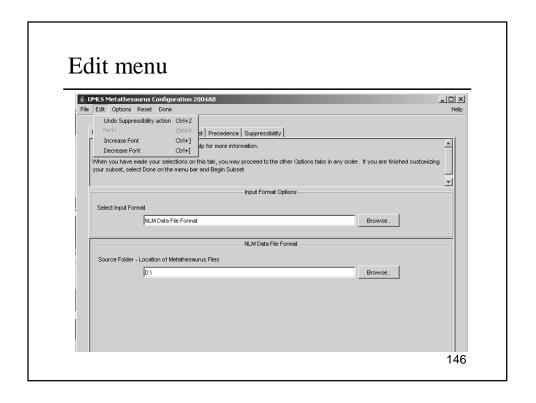


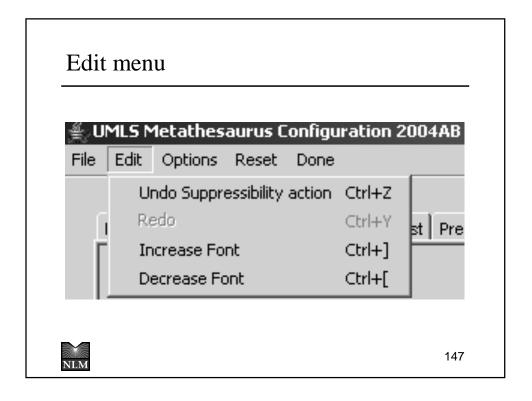


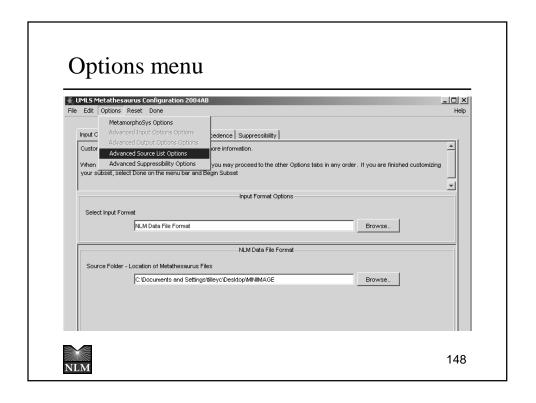


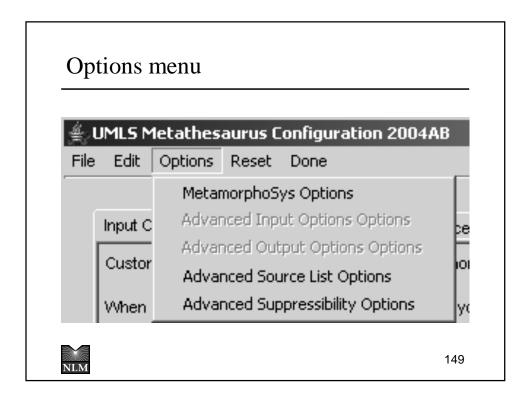


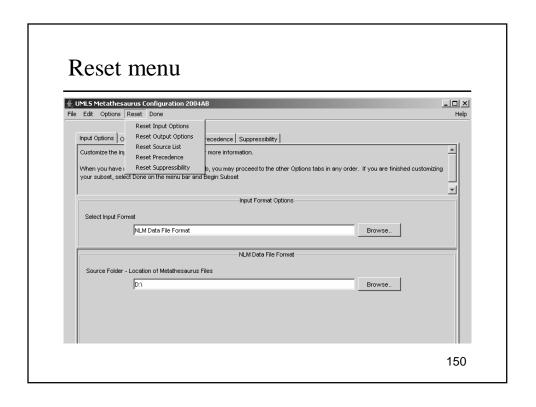




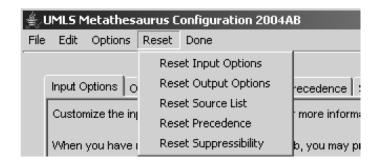






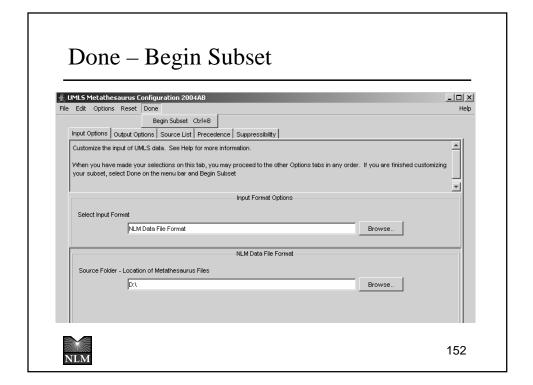


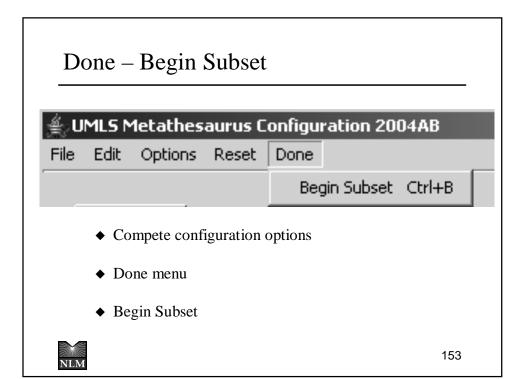
### Reset menu



- Returns all filters to default selections
- ◆ Default selections in "mmsys.prop.default file" in config folder
- mmsys.prop.default contains properties in last run







### How MetamorphoSys Works

- ◆ Removes all information from relational files in excluded vocabularies
  - atoms, strings, relationships, attributes, mappings, etc.
- ◆ Applies additional options selected by user
  - such as adding source term suppressibility or altering precedence
- ◆ Produces a full set of Metathesaurus files
  - relational files with customized data
  - reflecting other user criteria



### MetamorphoSys log



155

### MetamorphoSys log

```
mmsys.log - Notepad
  <u>File Edit Format View Help</u>
  Metathesaurus Output: Rich Release Format
 Long fields were not truncated.
Source Abbreviations were written out with a versionless (root) representation.
Fields containing UTF-8 characters were not removed.
 Excluded Sources
Kept Sources
AI/RHEUM, 1993
Alternative Billing Concepts
Alcohol and other Drug Thesaurus, 2000
Beth Israel Vocabulary, 1.0
Canonical Clinical Problem Statement System, 1999
Clinical Classifications Software, 2003
Current Dental Terminology (CDT), 4
COSTAR, 1989-1995
Medical Entities Dictionary, 2003
Physicians' Current Procedural Terminology, Spanish Translation,...
Physicians' Current Procedural Terminology, 2004
CRISP Thesaurus, 2004
COSTART, 1995
Diseases Database, 2000
German translation of ICD10, 1995
German translation of UMDNS, 1996

Kept Sources
                                                                                                                                                                                                   AIR93
                                                                                                                                                                                                   ALT2003
                                                                                                                                                                                                   AOD2000
                                                                                                                                                                                                   BI98
                                                                                                                                                                                                  CCPSS99
CCS2003
                                                                                                                                                                                                   COSTAR_89-95
                                                                                                                                                                                                   CPT01SP
                                                                                                                                                                                                   CPT2004
                                                                                                                                                                                                   CSP2004
CST95
                                                                                                                                                                                                   DDB00
                                                                                                                                                                                                   DMDICD10_1995
                                                                                                                                                                                                   DMDUMD_1996
```

Output directory contents								
N	F1	T						
Name	Size	Type 🔺						
(CHANGE)		File Folder						
indexes		File Folder						
፴ release.dat	1 KB	DAT File						
onfig.prop	8 KB	PROP File						
☑ AMBIGLUI.RRF	1,225 KB	RRF File						
	955 KB	RRF File						
MRCOC.RRF	809,207 KB	RRF File						
MRCOLS.RRF	21 KB	RRF File						
MRCONSO.RRF	596,528 KB	RRF File						
MRCUI.RRF	9,221 KB	RRF File						
MRCXT.RRF	9,391,778 KB	RRF File						
MRDEF.RRF	17,172 KB	RRF File						
MRDOC.RRF	88 KB	RRF File						
MRFILES.RRF	4 KB	RRF File						
MRHIER, RRF	899,786 KB	RRF File						
MRHIST.RRF	70,843 KB	RRF File						
MRMAP.RRF	9,362 KB	RRF File						

# Summary

### **UMLS** Overview

- ◆ UMLS = 3 Knowledge Sources
  - Metathesaurus
  - Semantic Network
  - SPECIALIST Lexicon and Lexical Tools
- ◆ MetamorphoSys
  - installs
  - customizes
- **◆** UMLSKS
  - remote access
  - resources and documentation



159

### Questions

### Documentation and Support

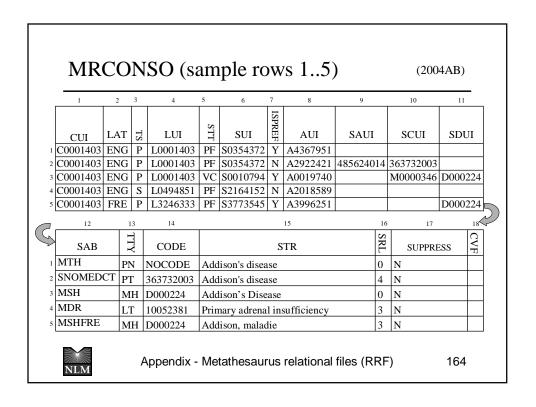
### UMLS documentation and support

- ◆ UMLS homepage http://umlsinfo.nlm.nih.gov/
  - with links to all other UMLS information
- ◆ UMLSKS homepage http://umlsks.nlm.nih.gov/
  - with links to the User's and Developer's guides
- ◆ Email address for support custserv@nlm.nih.gov

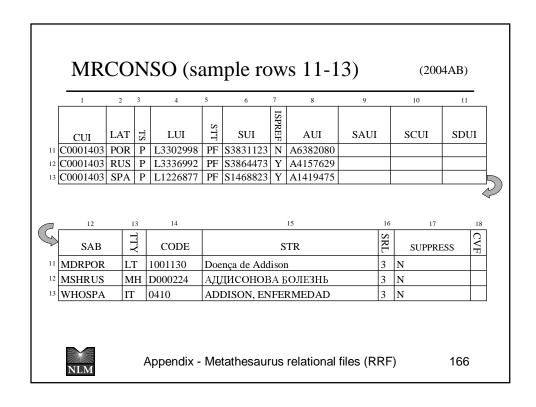


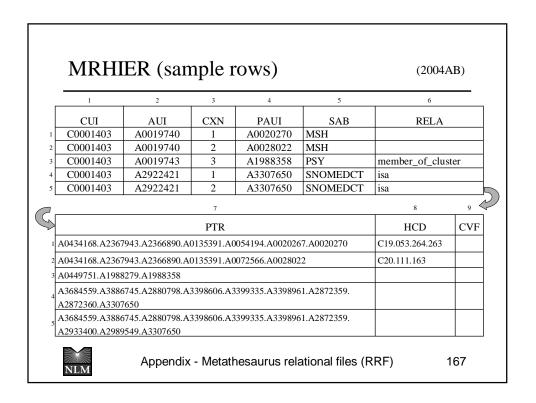
## Appendix 1

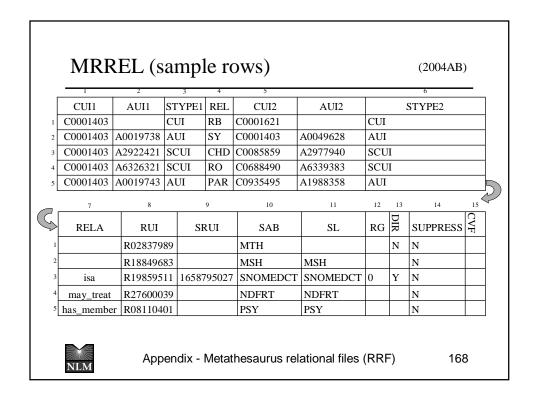
# UMLS files in Rich Release Format

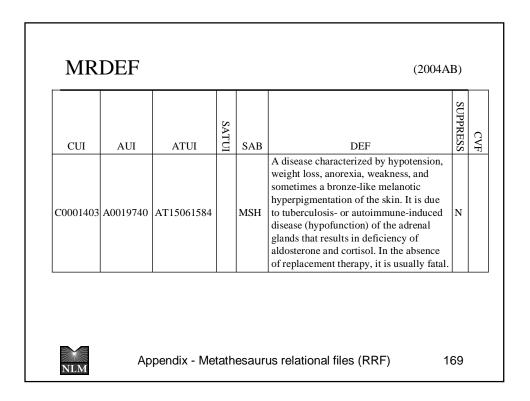


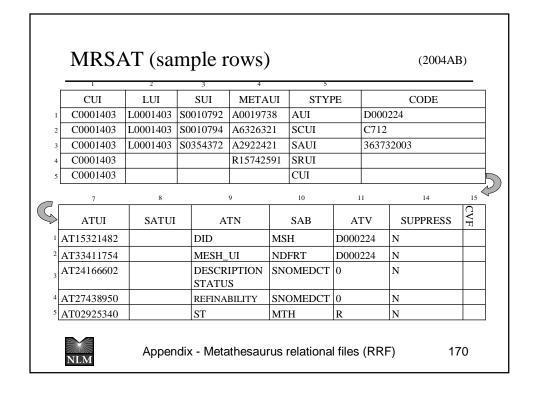
	MR	MRCONSO (sample rows 610)								(2004AB)		
	1	2	3	4	5	6	7	8	9		10	11
	CUI	LAT	TS	LUI	STT	SUI	ISPREF	AUI	SAUI		SCUI	SDUI
6	C0001403	FRE	S	L1272481	PF	S1514427	Y	A1464383				
7	C0001403	GER	P P	L1229627	PF	S1471573	Y	A4030156				D000224
8	C0001403	GER	S	L1239271	PF	S1481217	Y	A4034094				D000224
9	C0001403	JPN	P	L3437833	PF	S3965327	Y	A4264008				D000224
0	C0001403	JPN	S	L3465347	PF	S3992841	Y	A4291522				D000224
]	12		13	14				15		16	17	18
	SAB T CODE		STR				SRL	SUPPRI	ESS CY			
6	WHOFRE	ľ	Т	0410	MA	MALADIE D'ADDISON				2	N	
7	MSHGER MH D000224			Add	Addison-Krankheit			3	N			
8	MSHGER	MSHGER SY D000224 Bronzehautkrankheit				eit		3	N			
9	MSHJPN	N		D000224	Addison病				3	N		
0	MSHPJN	_		D000224	副腎性黒皮症 3			N				

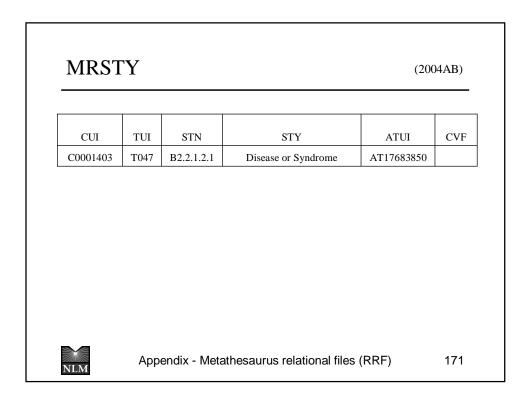


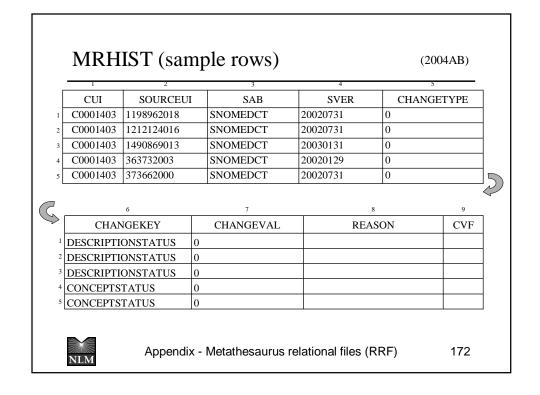












# Appendix 2

# UMLS files in Original Release Format

### MRCON Concepts

(2003AA)

```
LAT TS LUI
C0001403 ENG P L0001403 PF S0010794 Addison's Disease 0 C0001403 ENG P L0001403 VC S0352253 ADDISON'S DISEASE 0
C0001403 ENG P L0001403 VO S0010792 Addison Disease 0
C0001403 ENG P L0001403 VO S0033587 Disease, Addison 0
C0001403 ENG P L0001403 VO S0469271 Addison's disease, NOS 3 C0001403 ENG S L0278071 PF S0352321 ADRENAL INSUFFICIENCY (ADDISON'S DISEASE) 0
C0001403 ENG S L0278422 PF S0352329 ADRENOCORTICAL INSUFFICIENCY, PRIMARY FAILURE 0
C0001403 ENG | S | L0367999 | PF | S0469267 | Addison melanoderma | 3 |
C0001403 ENG | S | L0368000 | PF | S0496840 | Melasma addisonii | 3 | C0001403 | ENG | S | L0368398 | PF | S0506528 | Primary adrenal deficiency | 3 |
C0001403 ENG S L0373744 PF S0471237 Asthenia pigmentosa 3
C0001403 ENG | S | L0377831 | PF | S0473611 | Bronzed disease | 3 |
 \begin{array}{lll} \texttt{C0001403} \, | \, \texttt{ENG} \, | \, \texttt{S} \, | \, \texttt{L0494940} \, | \, \texttt{PF} \, | \, \texttt{S0718028} \, | \, \texttt{Primary} \, \, \, \texttt{adrenocortical} \, \, \texttt{insufficiency} \, | \, \texttt{3} \, | \, \\ \texttt{C0001403} \, | \, \texttt{ENG} \, | \, \texttt{S} \, | \, \texttt{L0494937} \, | \, \texttt{PF} \, | \, \texttt{S0718027} \, | \, \texttt{Primary} \, \, \, \texttt{adrenocortical} \, \, \, \texttt{insuff} \, | \, \texttt{3} \, | \, \\ \texttt{C0001403} \, | \, \texttt{ENG} \, | \, \texttt{S} \, | \, \texttt{L0494937} \, | \, \texttt{PF} \, | \, \texttt{S0718027} \, | \, \texttt{Primary} \, \, \, \texttt{adrenocortical} \, \, \, \texttt{insuff} \, | \, \texttt{3} \, | \, \texttt{1} \, | \, \texttt{
C0001403|FIN|P|L1510041|PF|S1805950|Addisonin tauti|3|
C0001403 FRE S L1272481 PF S1514427 MALADIE D'ADDISON 2 C0001403 GER P L1229627 PF S1471573 Addison-Krankheit 3
C0001403 GER S L1288823 PF S1530769 Primaere Nebennierenrindeninsuffizienz 1
C0001403 | ITA | P | L1276837 | PF | S1518783 | Morbo di Addison | 3 |
C0001403 POR P L0324623 PF S0432928 DOENCA DE ADDISON 2 C0001403 RUS P L0889403 PF S1093220 ADDISONOVA BOLEZN 3
C0001403 | SPA | P | L0342625 | PF | S0450930 | ENFERMEDAD DE ADDISON | 3 |
```

NLM

Appendix - Metathesaurus relational files (ORF)



### **MRSO** Sources

(2003AA)

LUI SUI SAB TTY SCD  ${\tt C0001403|L0001403|S0010792|MSH|EN|D000224|0|}$ C0001403 L0001403 S0010794 MSH MH D000224 0 C0001403 L0001403 S0010796 MSH PM D000224 0 C0001403 L0001403 S0010796 PSY PT 00810 3 C0001403 L0001403 S0033587 MSH PM D000224 0 C0001403 L0001403 S0220088 MSH PM D000224 0 C0001403 L0001403 S0352252 CCPSS PT 0022753 3 C0001403 L0001403 S0352252 DXP SY NOCODE 0 C0001403 L0001403 S0352253 CST GT ADREN INSUFFIC 0 C0001403 L0001403 S0352253 WHO IT 0410 2 C0001403 L0001403 S0354372 AOD DE 0000005430 0 C0001403 L0001403 S0354372 CSP PT 0060-3321 0 C0001403 L0001403 S0354372 LCH PT U000061 0 C0001403 L0001403 S0354372 MDR LT 10001130 3 C0001403 L0001403 S0354372 RCD PT C1541 3 C0001403 L0001403 S0354372 SNM SY D-2332 3 C0001403 L0001403 S0365923 CST GT ADREN INSUFFIC 0 C0001403 L0001403 S0469271 SNMI PT DB-70620 3 C0001403 | L0001403 | S1619433 | MDR | LT | 10001130 | 3 | C0001403 L0001403 S1911394 | ICPC2F | PT | T99002 | 3 | C0001403 | L0001403 | S1921523 | MTHICD9 | ET | 255.4 | 0 | C0001403 L0001403 S1932462 CPC2P SF T99002 3



Appendix - Metathesaurus relational files (ORF)

175



### MRDEF Definitions

(2003AA)

CUI SAB DEF

C0001403 MSH A disease characterized by hypotension, weight loss, anorexia, weakness, and sometimes a bronze-like melanotic hyperpigmentation of the skin. It is due to tuberculosis- or autoimmune-induced disease (hypofunction) of the adrenal glands that results in deficiency of aldosterone and cortisol. In the absence of replacement therapy, it is usually fatal.



Appendix - Metathesaurus relational files (ORF)



### MRSTY Semantic Types

(2003AA)

TUI STY C0001400 | T040 | Organism Function | C0001403 T047 Disease or Syndrome C0001406 T083 Geographic Area C0001407 T114 Nucleic Acid, Nucleoside, or Nucleotide C0001407 T123 Biologically Active Substance



Appendix - Metathesaurus relational files (ORF)



### MRATX Associated Expressions

SAB REL

Closed fracture of malar and maxillary bones, NOS C0009045|MSH|RB|<Zygomatic Fractures> OR <Maxillary Fractures> |

Unilateral congenital dislocation of hip

C0009702 MSH RB < Hip Dislocation, Congenital > AND < Femur Head > / < abnormalities > |

C0010700 | MSH | RB | <Bladder > / < surgery > |

C0010032 MSH RO Cornea>/<injuries>

CORRECTIVE LENS PROBLEM
C0010099|MSH|R0|<Contact Lenses>/<adverse effects>|

Chronic cough
C0010201|MSH|SY|<Cough> AND <Chronic Disease>|

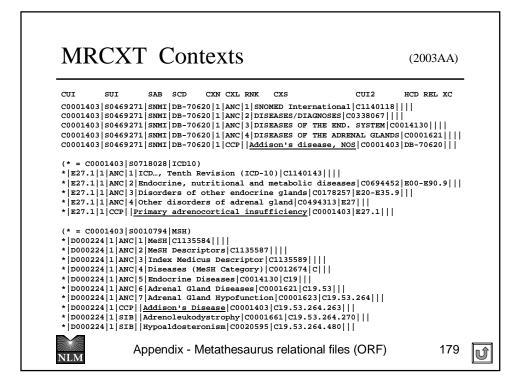
Cyst and pseudocyst of pancreas
C0010623|MSH|SY|<Pancreatic Cyst> OR <Pancreatic Pseudocyst>|

C0010692|LCH|RU|<Bladder>/<Inflammation>| ſ...1



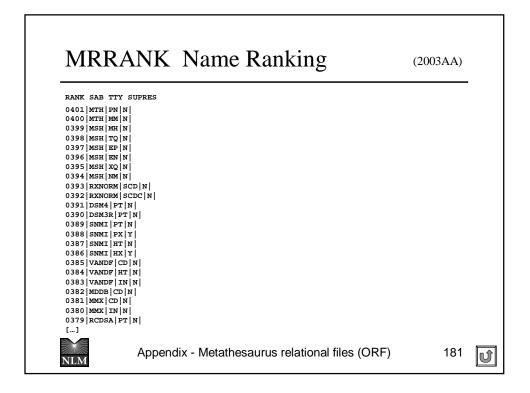
Appendix - Metathesaurus relational files (ORF)

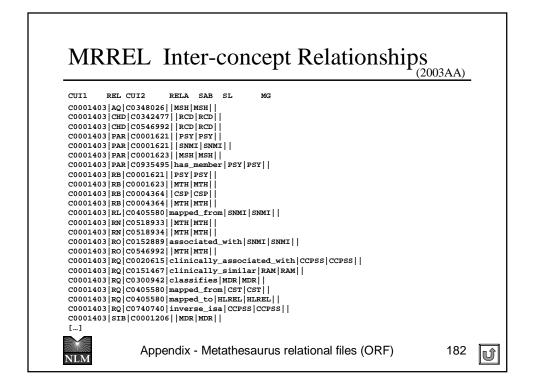


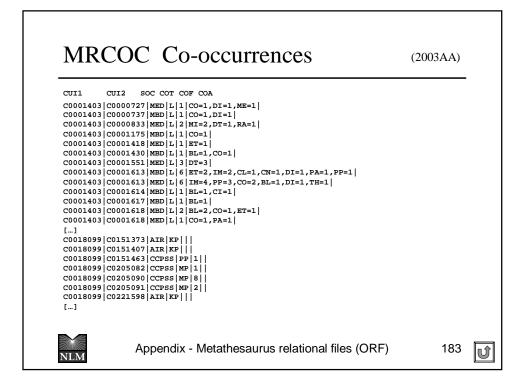


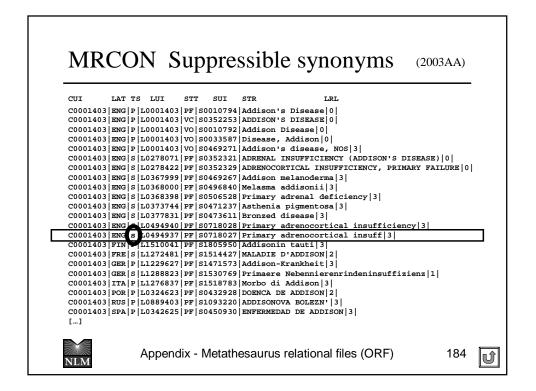
### MRSAT Simple concept attributes (2003AA)

```
ATN SAB
C0001403 | L0001403 | S0010792 | D000224 | DID | MSH | D000224 |
C0001403 L0001403 S0010792 D000224 EV MSH ADDISON DIS
C0001403 L0001403 S0010792 D000224 MUI MSH M0000346
C0001403 L0001403 S0010792 D000224 TH MSH UNK (19XX)
C0001403 L0001403 S0010794 D000224 AQL MSH BL CF CI CL CN CO DH DI DT EC EH EM EN ...
C0001403 L0001403 S0010794 D000224 DC MSH 1
C0001403 | L0001403 | S0010794 | D000224 | DID | MSH | D000224 |
C0001403 L0001403 S0010794 D000224 EV MSH ADDISON DIS C0001403 L0001403 S0010794 D000224 MDA MSH 19990101
C0001403 L0001403 S0010794 D000224 MED1963 NLM-MED
C0001403 | L0001403 | S0010794 | D000224 | MED1963 | NLM-MED | 2 |
C0001403 L0001403 S0010794 D000224 MED2002 NLM-MED *19
C0001403 L0001403 S0010794 D000224 MED2002 NLM-MED 23
C0001403|L0001403|S0010794|D000224|MN|MSH|C19.53.264.263|
C0001403 L0001403 S0010794 D000224 MN MSH C20.111.163
C0001403 L0001403 S0469271 DB-70620 SIC SNMI 255.4
C0001403||||DA|MTH|19900930|
C0001403||||MR|MTH|20021026|
C0001403||||ST|MTH|R|
               Appendix - Metathesaurus relational files (ORF)
```









### MRCUI Concept history

(2003AA)

CUI1 VER CREL CUI2 MAPIN
C0241779 | 1996AA | SY | C0001403 | Y |
C0271735 | 1996AA | SY | C0001403 | Y |
[...]



Appendix - Metathesaurus relational files (ORF)

185



### MRSAB Source information

(2003AA)

VCUI RCUI VSAB RSAB SON SF SVER MSTART MEND IMETA RMETA SLC SCC SRL TFR C1140103 | C1140104 | INS2002 | INS|French translation of the Medical Subject Headings, 2002 | MSH | 2002 | 2002 - 04\_11 | | 2002AB | | Dr. Annie Advocat; e-mail: advocat@inserm-dicdoc.u-strasbg.fr | Dr. Annie Advocat; e-mail: advocat@inserm-dicdoc.u-strasbg.fr | 3 | 30883 | 20692 | | MH,SY | | FRE | ISO646-US | Y | Y |

C1140132|C1140133|BRMP2002|BRMP|Portuguese translation of the Medical Subject Headings, 2002|MSH|2002|2001\_12\_04||2002AA||Elenice de Castro; e-mail: elenice@brm.bireme.br|Elenice de Castro; e-mail: elenice@brm.bireme.br|Slenicedecastro; e-mail: elenice@brm.bireme.br|3|41853|27195||EP,MH,SY||POR|ISO646-US|Y|Y|

C1140297 | C1140298 | DUT2001 | DUT | Dutch Translation of the Medical Subject Headings, 2001 | MSH | 2001 | 2001 | 12 - 04 | | 2002AB | | A.J.P.M.Overbeke, overbeke@ntvg.nl, \* 20 662 0150 | A.J.P.M.Overbeke, overbeke@ntvg.nl, \* 20 662 0150 | 3 | 35705 | 17733 | | EP,MH,SY | | DUT | ISO646-US | Y | Y |



Appendix - Metathesaurus relational files (ORF)



#### SRDEF Basic information

(2003AA)

```
RT TUI STY/RL STN/RTN DEF
                                                                                                                   UN
                                                                                                                                                          ABR
                                                                                                                                                                                 RIN
{\tt STY} \\ | \\ {\tt T001} \\ | \\ {\tt Organism} \\ | \\ {\tt A1.1} \\ | \\ {\tt Generally, a living individual, including all plants and of the property o
animals. | Homozygote; Radiation Chimera; Sporocyst | | | | |
STY | T002 | Plant | A1.1.1 | An organism having cellulose cell walls, growing by
synthesis of inorganic substances, generally distinguished by the presence of
chlorophyll, and lacking the power of locomotion. Plant parts are included here
as well. |Pollen; Potatoes; Vegetables | | | | |
STY T003 Alga Al.1.1.1 A chiefly aquatic plant that contains chlorophyll, but does
not form embryos during development and lacks vascular tissue. |Chlorella;
Laminaria; Seaweed | | | | |
\mathtt{STY} \, | \, \mathtt{T004} \, | \, \mathtt{Fungus} \, | \, \mathtt{A1.1.2} \, | \, \mathtt{A} \, \text{ eukaryotic organism characterized by the absence of}
chlorophyll and the presence of a rigid cell wall. Included here are both slime molds and true fungi such as yeasts, molds, mildews, and mushrooms. Aspergillus
clavatus; Blastomyces; Helminthosporium; Neurospora | | | | |
RL|T132|physically_related_to|R1|Related by virtue of some physical attribute or
 characteristic.||||PR|physically_related_to|
RL|T133|part_of|R1.1| Composes, with one or more other physical units, some larger whole. This includes component of, division of, portion of, fragment of, section
of, and layer of. | | | PT | has_part |
RL T186 isa H The basic hierarchical link in the Network. If one item "isa"
another item then the first item is more specific in meaning than the second
item.||||IS|inverse_isa|
ſ...1
```



Appendix - Semantic Network relational files (ORF)

187



### SRSTR Structure

(2003AA)

```
STY/RL
                                STY/RL
{\tt Biologic\ Function} | {\tt affects} | {\tt Organism} | {\tt D} |
Biologic Function isa Natural Phenomenon or Process D
Biologic Function process_of Organism D
Biologic Function produces Biologically Active Substance D
Biologic Function produces Body Substance D
[...]
Disease or Syndrome | conceptually_related_to | Experimental Model of Disease | DNI |
 \begin{tabular}{ll} Disease or Syndrome is a Pathologic Function | D | \\ \end{tabular} 
Disease or Syndrome|produces|Tissue|D|
[...]
{\tt Medical\ Device | isa | Manufactured\ Object | D |}
Medical Device prevents | Injury or Poisoning | D |
Medical Device prevents Pathologic Function D
Medical Device treats Anatomical Abnormality D
Medical Device treats Injury or Poisoning D
Medical Device | treats | Pathologic Function | D |
Medical Device treats Sign or Symptom D
[...]
{\tt Mental\ Process[process\_of[Plant]B|\ blocks\ Biologic\ Function[process\_of]Organism[D]}
[...]
part_of | isa | physically_related_to | D |
```



Appendix - Semantic Network relational files (ORF)



